Individual tree-crown detection in RGB imagery using self-supervised deep learning neural networks

Ben. G. Weinstein[1], Sergio Marconi[1], Stephanie Bohlman[2], Alina Zare[3], Ethan White[1]

[1] Department of Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida, USA

[2] School of Forest Resources and Conservation, University of Florida, Gainesville, Florida, USA

[3] Department of Electrical and Computer Engineering, University of Florida, Gainesville, Florida, USA

Abstract

Remote sensing can transform the speed, scale, and cost of biodiversity and forestry surveys. Data acquisition currently outpaces the ability to identify individual organisms in high resolution imagery. We outline an approach for identifying tree-crowns in true color, or red/green blue (RGB) imagery using a deep learning detection network. Individual crown delineation is a persistent challenge in studies of forested ecosystems and has primarily been addressed using three-dimensional LIDAR. We show that deep learning models can leverage existing lidar-based unsupervised delineation approaches to initially train an RGB crown detection model, which is then refined using a small number of hand-annotated RGB images. We validate our proposed approach using an open-canopy site in the National Ecological Observation Network (NEON). Our results show that combining LIDAR and RGB methods in a self-supervised model improves predictions of trees in natural landscapes. The addition of a small number of hand-annotated trees improved performance over the initial self-supervised model. While undercounting of individual trees in complex canopies remains an area of development, deep learning can increase the performance of remotely sensed tree surveys.

Keywords: Deep Learning; Trees; Detection; Remote Sensing; LIDAR; RGB; NEON

1. Introduction

The costs of human observations of biological phenomena limit our ability to understand the natural world. By embracing image-based artificial intelligence, biology can advance our understanding of individual organisms, species, and ecosystems (Anderson, 2018). The growing availability of sub-meter airborne imagery brings opportunities for remote sensing of biological landscapes that scales from individual organisms to global systems. The remaining hurdle is the move from laborious, non-reproducible, and costly annotation of these datasets to automated, reproducible extraction of biological information (Weinstein, 2018).

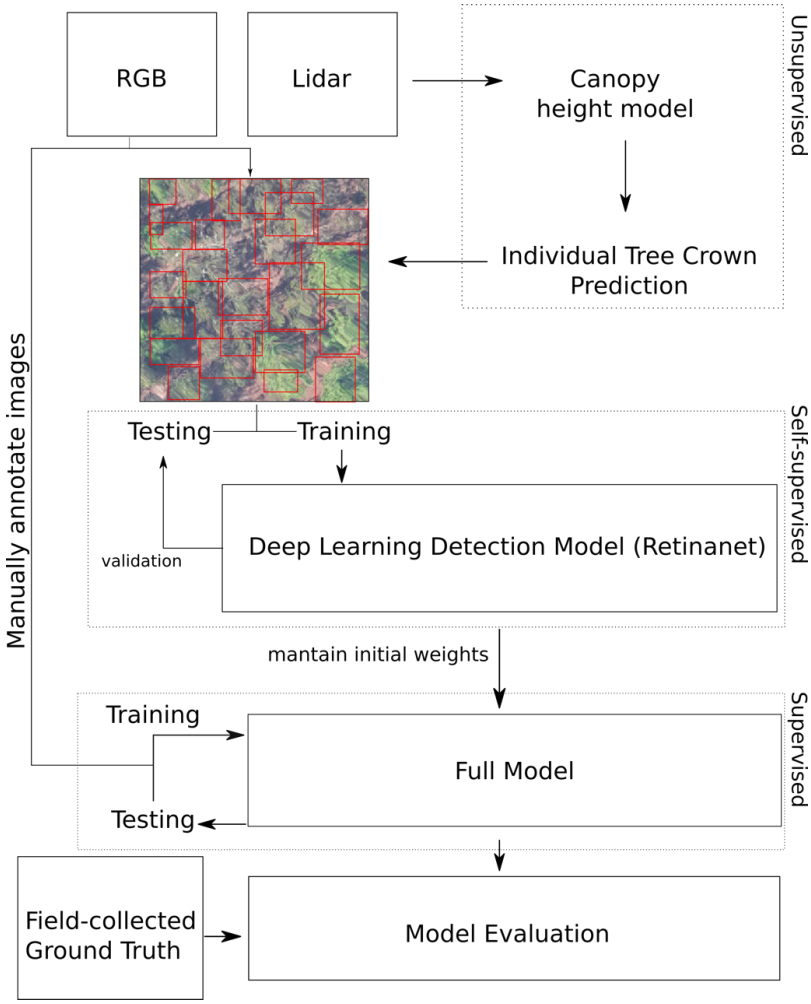
Tree detection is a central task in forestry and ecosystem research, and both commercial and scientific applications rely on delineating individual tree crowns from imagery (Caughlin et al., 2016; Wu et al., 2016). While there has been considerable research in tree detection using lidar-based unsupervised classification (Ayrey et al., 2017; Liu et al., 2015; Wu et al., 2016), less is known about supervised tree detection in RGB orthophotos. Compared to LIDAR, two dimensional RGB is less inexpensive to acquire, easy to process, but lacks three-dimensional information on crown shape. In addition, RGB data has a long historical record, whereas widespread LIDAR is a recent development. Effective RGB-based tree detection would unlock data at much larger scales due to increasing satellite-based RGB resolution and the growing use of unmanned aerial vehicles.

The promise of deep learning for airborne biodiversity detection is three-fold. First, convolutional neural networks (CNNs) learn from training data, rather using hand-crafted pixel features, to delineate objects of interest. This reduces the expertise required for each use-case and improves transferability among projects (Ayrey and Hayes, 2018). Second, CNNs learn

hierarchical combinations of image features, thereby reducing the reliance on individual pixels, which vary due to the acquisition environment. Finally, neural networks are re-trainable to incorporate the idiosyncrasies of individual datasets. This means that models can be refined with data from new local areas, without discarding information from previous training sets.

The challenge for applying deep learning to natural systems is the need for large training datasets. The quality and quantity of training data impacts both prediction accuracy and transferability. Collecting sufficient training data is expensive and logistically difficult. The high variation in tree crown appearance due to taxonomy, health status, and human intervention increases the risk of overfitting when using small amounts of training data (Li et al., 2016). More broadly, a lack of sufficient training is a pervasive problem in machine learning of remotely sensed imagery (Zhu et al., 2017). To address this challenge, recent approaches have generated training data from unsupervised classification algorithms (Wu and Prasad, 2018). The output of the unsupervised classification is then used to train a supervised model. We refer to this as *self-supervision*, due to the unsupervised generation of training data. This initial training provides important regularization for the network, even though the labeled data are imperfect due to the limitations of the unsupervised classification algorithm (Erhan et al., 2009). This initial training is followed by retraining using a small number of hand-annotations to correct errors from the unsupervised classification. We implemented this workflow using a LIDAR unsupervised classification to generate training trees for RGB supervised learning (Scheme 1), and then amended the initial training set with hand-annotated trees. The LIDAR data is solely used to bolster the initial training of the network but is not used for the final training step. The

result is a deep learning neural network that can perform tree delineation in new RGB imagery without the need for co-registered LIDAR data.



Scheme 1. A conceptual figure of the proposed pipeline. A LIDAR-based unsupervised classification generates initial training data for a self-supervised RGB deep learning model. The model is then retrained based on a small number of hand-annotated trees to create the full model.

1.1. Relation to previous work

Initial studies of tree detection in RGB imagery focused on pixel-based methods and watershed algorithms to find local maxima among pixels to create potential tree crowns (e.g. Gougeon and Leckie, 2006). Combined with hand-crafted rules on tree geometries, these approaches separately performed tree-detection and crown delineation (Ke and Quackenbush, 2011). More recently developed growing region algorithms focus on lidar-based parameterized models of tree shape to simultaneously detect trees and establish crown boundaries (Gomes et al., 2018; Weinmann et al., 2017). These approaches are limited by the need to choose parameters that encompass a variety of tree forms. For example, Coomes et al. (2017) showed that adding allometric relationships between trunk height and crown width improved lidar-based segmentation, but the optimal relationship will vary based on tree species, age class, and biotic neighborhood. This makes creating a single set of rules that encompass the range of tree types challenging (Yin and Wang, 2016).

Deep learning has only been recently applied to airborne forestry measurements (Ayrey and Hayes, 2018; Li et al., 2016). After delineating trees using LIDAR, several papers have used convolutional neural networks to assign species labels to candidate trees (Deng et al., 2016; Mizoguchi et al., 2017). To our knowledge, the only prior use of deep learning for vegetation detection showed that passing a sliding window CNN over the entire image outperformed pixel-based watershed methods for detecting individual trees in palm plantations (Li et al., 2016) and small shrubs in arid landscapes (Guirado et al., 2017). Our study advances this research by using

advances in region proposal networks and a self-supervised approach to alleviate the issues with limited labeled data.

2. Materials and Methods

2.1. Study Site and Field Data

We used data from the National Ecological Observatory Network (NEON) site at the San Joaquin Experimental Range in California to assess our proposed approach (Figure 1). The site contains open woodland of live oak (*Quercus agrifolia*), blue oak (*Quercus douglasii*) and foothill pine (*Pinus sabiniana*) forest. The majority of the site is a single-story canopy with mixed understory of herbaceous vegetation. All aerial remote sensing data products were provided by the NEON Airborne Observation Platform. We used the NEON 2018 "classified LiDAR point cloud" data product (NEON ID: DP1.30003.001), and the "orthorectified camera mosaic" (NEON ID: DP1.30010.001). The LiDAR data consist of 3D spatial point coordinates (4-6 points/m^2) which provides high resolution information about crown shape and height. The RGB data are a 1km x 1km mosaic of individual images, with a cell size of 0.1m meters. Both data products are georeferenced in the UTM projection Zone 11. In addition to airborne data, NEON field teams semi-annually catalog "Woody Plant Vegetation Structure" (NEON ID: DP1.10098.001), which lists the tag and species identity of trees with DBH > 10cm in 40m x 40m plots at the site. For each tagged tree, the trunk location was obtained using the azimuth and distance to the nearest georeferenced point within the plot. All data are publicly available on the NEON Data Portal (http://data.neonscience.org/). All code for this project is available in a GitHub repository (https://github.com/weecology/DeepForest) (Weinstein and White 2019).

For hand annotations, we selected a random 1km x 1km RGB tile and used the program RectLabel (https://rectlabel.com/) to draw bounding boxes around each tree. We chose not to include snags, or low bushes that appeared to be non-woody. In total, we annotated 1988 trees for the San Joaquin site. In addition to the 1km tile, we hand drew the canopy boxes on the cropped RGB images for each NEON plot (n=35), which were withheld from training and used as a validation dataset.

Figure 1. Each NEON site is covered by a mosaic of high-resolution 1km^2 tiles. A sample tile from San Joaquin, CA shows the high-resolution RGB images needed for tree detection and classification in airborne imagery.

2.2. Unsupervised LIDAR Classification

The first step in our pipeline is the use of an existing unsupervised classification algorithm from Silva et al. (2015) to create initial tree predictions in the LIDAR point cloud. This algorithm uses a canopy height model and threshold of tree height to crown width to cluster the LIDAR cloud into individual trees. We used a canopy height model of 0.5m resolution to generate local tree tops, and a maximum crown diameter of 60% of tree height. A bounding box was automatically drawn over the entire set of points assigned to each tree to create the tree prediction. Example results of the LIDAR derived unsupervised classification are shown in Figure 2a,b and Figure 5a.

2.3. Deep Learning RGB detection

Convolutional neural networks are often used for object detection, due to their ability to represent semantic information as combinations of image features. Early applications passed a sliding window over the entire image, classifying each window as a foreground-background class. This approach was slow and enforced arbitrary decisions for window size and shape. This was improved on by two-stage proposal networks that used either an image segmentation proposal system (Uijlings et al., 2013) or a region-proposal network to generate diverse bounding boxes based on image features (Ren et al., 2015). Recently, one-stage detectors have increased the speed of object detection by combining the regional proposal and classification into a single workflow. We chose the Retinanet one-stage detector (Lin et al., 2017), which has two additions to previous one stage region-proposal detection networks. The first is a set of hierarchical feature pyramids that merge information from different scales. This cross-scale learning is critical for objects, such as trees, that vary in size. The second contribution is focal loss, which minimizes the foreground-background class imbalance that is common in dense

sampling of one-stage detection networks. Since the majority of anchors will not contain a foreground object, the focal loss down-weights the importance of easily predicted boxes, thereby reducing the effect on model weights. The result is a fast detection network that has shown strong performance in traditional computer vision benchmarks and ecological applications (Levy et al., 2018). We used a resnet-50 classification backbone pretrained on the ImageNet dataset (He et al., 2016). We experimented with deeper architectures (resnet-101 and resnet-152) but found no improvement that offset the increased training time.

Since the entire 1km RGB tile cannot fit into GPU memory, we first cut the tile into smaller windows for model training. We experimented with a number of different window sizes and found optimal performance at 400 X 400 pixels due to a balance between memory constraints and providing the model sufficient spatial context for tree detection. This resulted in 729 windows per 1km tile. The order of tiles and windows were randomized before training. From the pool of unsupervised tree predictions, we selected 20,000 windows and trained with a batch size of 6 on a Tesla K80 GPU for 10 epochs. After prediction, we passed each image through a non-max suppression filter to remove predicted boxes that overlapped by more than 15%. In addition, one advantage of the neural network approach is that each predicted bounding box has an associated confidence score. We removed boxes within confidence scores less than 0.15.

2.4. Model Evaluation

We used the NEON woody vegetation data to evaluate model recall using field-collected points corresponding to individual tree trunks. A field-collected tree point was considered correctly predicted if the point fell within a predicted bounding box. To evaluate model precision, we

used the mean average precision (mAP) score for the hand-annotated datasets. The mAP metric is a summary of the average precision across a range of recall values. To compute this metric, we sort predicted boxes by their confidence score, and then selects the top k boxes, where k is the number of ground truth samples in the image. For each of k boxes, a ground truth sample is considered correctly predicted if it has an intersection-over-union score of greater than 0.5 (referred to as mAP @50). The intersection over union evaluation metric measures the area of overlap divided by the area of union of the ground truth bounding box and the predicted bounding box.

3. Results

The proposed pipeline predicted more than 88% of the field collected tree points, with a mAP@50 precision of 0.50 for the hand-annotated validation data (Figure 2). The full model performed better than each of the component parts, with increases in both recall and precision using a combination of pre-training on lidar-based unsupervised data and a small number of hand-annotated tree crowns (Table 1).

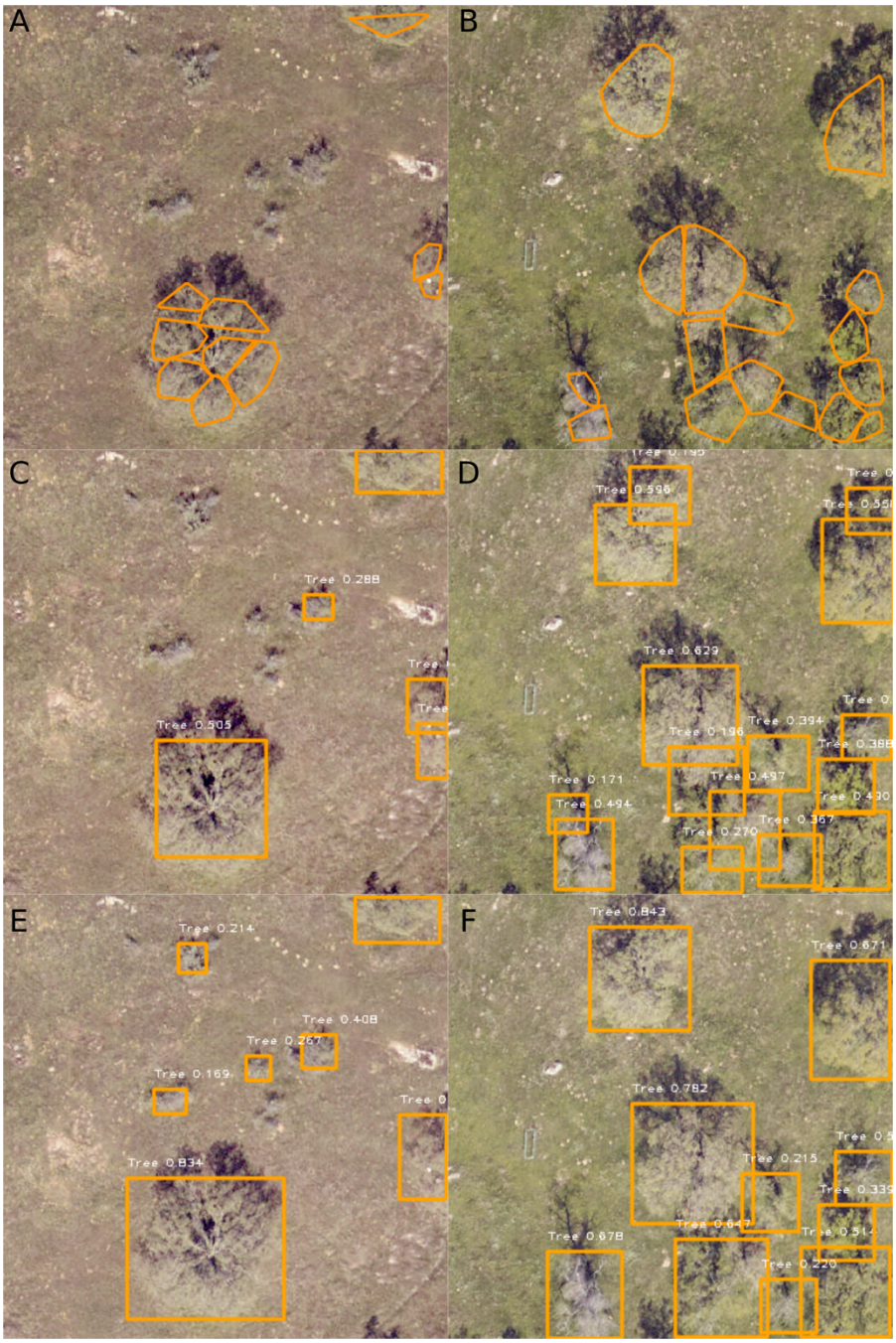


Figure 2. Predicted individual tree crowns for the unsupervised lidar (A, B), self-supervised RGB (C, D) and full model (E, F) for two NEON tower plots, SJER_015 (A, C, E), and SJER_053 (B, D, F) at the San Joaquin, CA site. For each tree prediction, the detection probability is shown in white.

After each tile had been preprocessed and split into the 400x400 windows, prediction performance was 1 image per 5.19 seconds on CPU and 1 image per 0.2 seconds on GPU. This translates to a 1km x 1km tile in 2.43 GPU minutes, with the entire NEON site in six hours (150 tiles). In comparison, the unsupervised lidar classification took 4.46 minutes per 1km tile. We cannot report a comparable mAP score for the unsupervised LIDAR classification, because it does not provide confidence scores. However, we can report that only 7.6% of the polygons predicted had intersection-over-union of greater than 50% with the hand-annotated boxes due to consistent over-segmentation of large trees. This value should not be seen as direct comparison to the RGB mAP scores. In addition, it should be noted that our quantitative results are likely biased toward the RGB model, since the hand-annotations were made by looking at the RGB, and not the LIDAR data.

Table 1. Evaluation metrics for each of the candidate models. Recall was calculated using the field-collected tree points from the NEON tower plots (n=34). Precision was calculated on hand annotated bounding boxes around tree crown for the 34 NEON tower plots (n=271 trees). These The unsupervised LIDAR algorithm does not compute probability scores, it is therefore not possible to report the mAP metric for this model.

Model	**Precision (mAP@50)**	**Recall**
Unsupervised LIDAR	-	0.65
Hand-annotation only	0.34	0.84
Self-supervised RGB	0.26	0.83
Full Model (self-supervised + hand-annotation)	0.50	0.88

By reviewing images predicted by the unsupervised lidar classification, the self-supervised RGB deep learning model, and the full model, we can learn about the contributions of each stage of the pipeline. The LIDAR unsupervised classification does a good job of identifying trees versus background based on height. Most small trees are well segmented, but there is consistent over-segmentation of the large trees, with multiple crown predictions abutting together. Visual inspection shows that these predictions represent multiple major branches of a single large tree, rather than multiple small trees (Figure 2a). In the self-supervised RGB model, these large trees are more accurately segmented, but there is a proliferation of bounding boxes, and overall lower confidence scores for even well-resolved trees (Figure 2d). This is evident in the precision-recall curves for the hand-annotated validation data, in which the self-supervised model more rapidly declines in performance at higher score thresholds (Figure 3). By combining the self-supervised and the hand annotated datasets, the full model reduces the extraneous boxes and improves the segmentation of large trees (Figure 4). The full model has optimal performance in areas of well-spaced large trees (Figure 4b) but tends to under-segment small clusters of trees (Figure 4c).

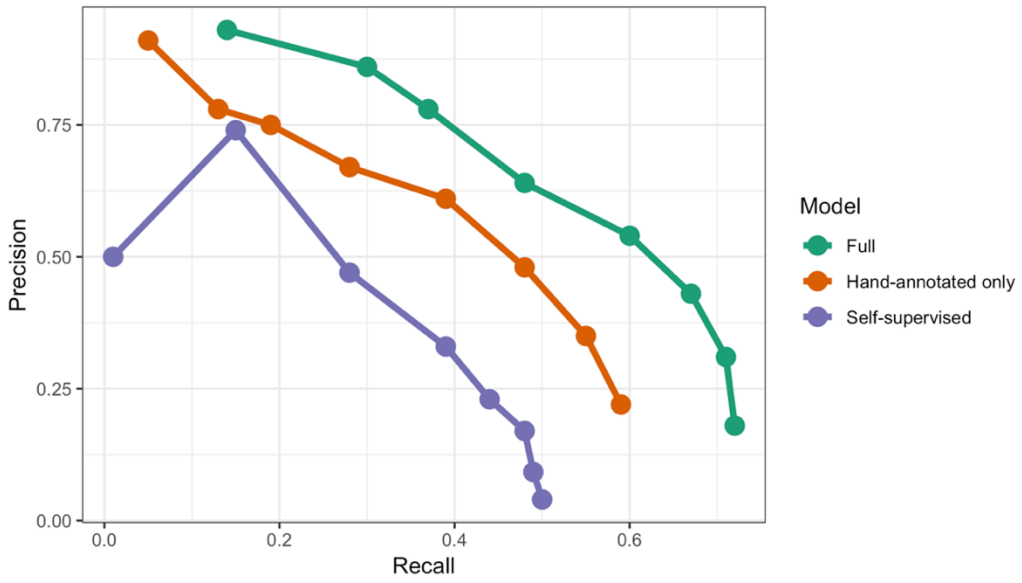


Figure 3. Precision-recall curves for the hand-annotated NEON plots. For each model, we calculated the proportion of correctly predicted boxes for score thresholds [0,0.1,..,0.7]. An annotation was considered correctly predicted if the intersection-over-union (IoU) score was greater than 0.5. Note that the recall on the x-axis corresponds to the proportion of true positives in the hand-annotated data, and not the field-collected centroids in Table 1.

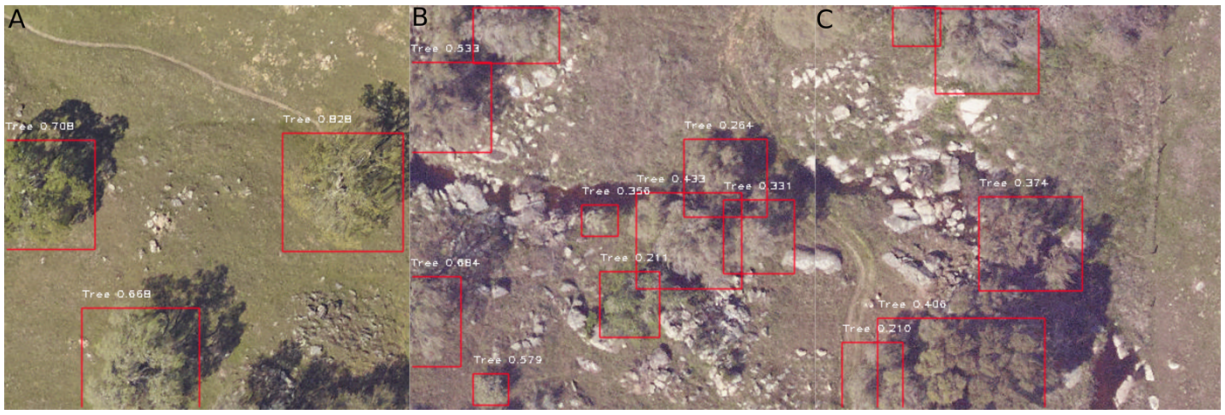


Figure 4. Predictions from the full model on the validation 1km x 1km tile. Canopy complexity increases from a) well-defined large trees to B) mixed-species canopies, c) tightly packed clusters of trees. As canopy complexity increases, the full model tends to under-segment small tree clusters.

4. Discussion

Using recent developments in deep learning we built a neural network-based pipeline for identifying individual trees in RGB imagery. Commercial high resolution RGB data is increasingly available at near global scales, meaning that an accurate RGB based crown delineation methods could be used to detect overstory trees at unprecedented extents. To address the long-standing challenge of a lack of labeled training data, we used an unsupervised LIDAR classification to generate labels for initial training. This self-supervised approach allows the network to learn the general features of trees even if the LIDAR-based unsupervised classification is imperfect. The addition of only 2,000 hand-annotated trees generated a final model that performed well when applied to a large geographic area. This approach opens the door for the use of deep learning in airborne biodiversity surveys, despite the persistent lack of annotated data in forestry and ecology datasets.

While our method uses LIDAR data to train the initial RGB model, it is not needed for prediction. This means that once trained, the model can be deployed anywhere that high quality RGB data exists. An unexpected benefit of the RGB model was the ability to discriminate trees from other vertical objects, such as houses or poles, despite a lack of distinction in the unsupervised LIDAR training data (Figure 5). This will be useful in urban tree detection and other non-forested sites.

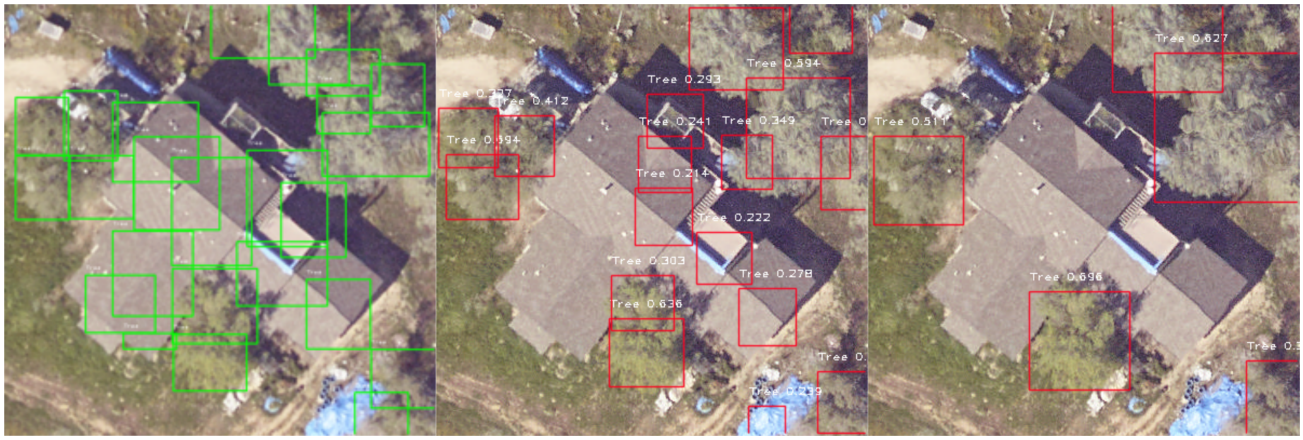


Figure 5. Improvement in prediction quality during the training pipeline. A) Ground truth based on the lidar-based unsupervised classification erroneously segmented man-made structures as trees. B) Predictions from the self-supervised RGB model showed that the addition of RGB data diminished the effect of incorrect labeled training data, with only edges of the man-made structure maintained as tree predictions. C) Combining the self-supervised RGB data with hand-annotations eliminated the influence of the original misclassification in the training data, while still capturing the majority of trees in the image.

In general, it is likely that accurate tree detection will be region specific, and that the best model will vary among environments. This will require training a new model for a general region, using both RGB and LIDAR training data. Our approach should save resources by allowing a smaller scale LIDAR flight to generate training data, and then cover a much larger area with inexpensive RGB orthophotos. The permanent 45 NEON plots were selected to cover of common ecological domains and could therefore serve as pools of LIDAR and RGB data for regional model training. Combining these detectors together could produce tree detection maps at broad scales, with potential applications to ecosystem health, post-natural disaster recovery, and carbon dynamics.

Both RGB and LIDAR data capture information useful for tree-detection, and our results show increased performance when used together. Compared to the unsupervised LIDAR classification, the deep learning model more closely resemble hand-annotated trees. One remaining challenge is providing a definition for what is a tree versus smaller statured vegetation such as shrubs. For example, small trees were often considered too low for inclusion in the LIDAR algorithm (Figure 2a), whereas they were included in the full model based on the hand-annotations (Figure 2b). When deploying these models to applied problems, it will be important to have strict quantitative guidelines that define class definitions.

While our method performed well at our open-canopy test site, we anticipate that geographic areas with complex canopy conditions will be more challenging. The current model solely uses LIDAR in the pretraining step. Where available, directly incorporating a LIDAR canopy height model into the deep learning approach will allow the model to simultaneously learn the vertical features of individual trees in addition to their two-dimensional color features in the RGB data. Recent applications of three-dimensional CNNs (Zhou and Tuzel, 2017), as well as point-based semantic segmentation (Qi et al., 2017), provide new avenues for joint multi-sensor modeling. These developments will be crucial in segmenting complex canopies that overlap in the two-dimensional RGB imagery. In addition, recent extensions of region-proposal networks refine bounding boxes to identify the individual pixels that belong to a class (He et al., 2017). This will provide a better estimate of tree crown area, as trees often have a non-rectangular shape.

5. Conclusions

Applying deep learning models to natural landscapes opens new opportunities in ecology, forestry, and land management. In addition to scaling tree detection at much lower costs, there is the potential to provide additional important information about natural systems. The current model could be expanded from a single class, “Tree”, to one that provides more detailed classifications based on taxonomy and health status. For example, splitting the “Tree” class into living and dead trees would provide management insight when surveying for outbreaks of tree pests and pathogens (Wulder et al., 2006), as well as post-fire timber operations (Vogeler et al., 2016). With the addition of hyperspectral data, dividing the tree class into species labels yields additional insights into the economic value, ecological habitat, and carbon storage capacity for large geographic areas (Deng et al., 2016). As such, deep learning-based approaches provide the potential for large scale actionable information on natural systems to be derived from remote sensing data.

6. Author Contributions

BGW, EPW, SB and AZ conceived of project design. EW and SM collected the preliminary data. BGW performed the analysis and wrote the text. All authors contributed to the text.

7. Funding

This research was supported by the Gordon and Betty Moore Foundation’s Data-Driven Discovery Initiative through grant GBMF4563 to E.P. White. The authors declare no conflict of interest.

8. References

Anderson, C.B., 2018. Biodiversity monitoring, earth observations and the ecology of scale. Ecol.

Lett. https://doi.org/10.1111/ele.13106

Ayrey, E., Fraver, S., Kershaw, J.A., Kenefic, L.S., Hayes, D., Weiskittel, A.R., Roth, B.E., 2017. Layer Stacking: A Novel Algorithm for Individual Forest Tree Segmentation from LiDAR Point Clouds. Can. J. Remote Sens. 43, 16–27. https://doi.org/10.1080/07038992.2017.1252907

Ayrey, E., Hayes, D., 2018. The Use of Three-Dimensional Convolutional Neural Networks to Interpret LiDAR for Forest Inventory. Remote Sens. 10, 649. https://doi.org/10.3390/rs10040649

Caughlin, T.T., Graves, S.J., Asner, G.P., Van Breugel, M., Hall, J.S., Martin, R.E., Ashton, M.S., Bohlman, S.A., 2016. A hyperspectral image can predict tropical tree growth rates in single-species stands. Ecol. Appl. 26, 2367–2373. https://doi.org/10.1002/eap.1436

Coomes, D.A., Dalponte, M., Jucker, T., Asner, G.P., Banin, L.F., Burslem, D.F.R.P., Lewis, S.L., Nilus, R., Phillips, O.L., Phua, M.H., Qie, L., 2017. Area-based vs tree-centric approaches to mapping forest carbon in Southeast Asian forests from airborne laser scanning data. Remote Sens. Environ. 194, 77–88. https://doi.org/10.1016/j.rse.2017.03.017

Deng, S., Katoh, M., Yu, X., Hyyppä, J., Gao, T., 2016. Comparison of tree species classifications at the individual tree level by combining ALS data and RGB images using different algorithms. Remote Sens. 8. https://doi.org/10.3390/rs8121034

Erhan, D., Manzagol, P.-A., Bengio, Y., Bengio, S., Vincent, P., 2009. The Difficulty of Training Deep Architectures and the Effect of Unsupervised Pre-Training. Twelfth Int. Conf. Artif. Intell. Stat. (AISTATS), JMLR Work. Conf. Procedings 5, 153–160. https://doi.org/10.1016/j.jhealeco.2010.06.004

Gomes, M.F., Maillard, P., Deng, H., 2018. Individual tree crown detection in sub-meter satellite imagery using Marked Point Processes and a geometrical-optical model. Remote Sens. Environ. 211, 184–195. https://doi.org/10.1016/j.rse.2018.04.002

Gougeon, F.A., Leckie, D.G., 2006. The Individual Tree Crown Approach Applied to Ikonos Images of a Coniferous Plantation Area. Photogramm. Eng. Remote Sens. https://doi.org/10.14358/PERS.72.11.1287

Guirado, E., Tabik, S., Alcaraz-Segura, D., Cabello, J., Herrera, F., 2017. Deep-learning Versus OBIA for Scattered Shrub Detection with Google Earth Imagery: Ziziphus lotus as Case Study. Remote Sens. 9, 1220. https://doi.org/10.3390/rs9121220

He, K., Gkioxari, G., Dollar, P., Girshick, R., 2017. Mask R-CNN. Proc. IEEE Int. Conf. Comput. Vis. 2017–Octob, 2980–2988. https://doi.org/10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. Comput. Vis. Pattern Recognit. (CVPR), 2016 770–778. https://doi.org/10.3389/fpsyg.2013.00124

Ke, Y., Quackenbush, L.J., 2011. A review of methods for automatic individual tree-crown detection. Int. J. Remote Sens. 32, 4725–4747.

Levy, D., Belfer, Y., Osherov, E., Bigal, E., Scheinin, A.P., Nativ, H., Tchernov, D., Treibitz, T., King, A., Bhandarkar, S.M., 2018. Automated Analysis of Marine Video With Limited Data. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Work. 1385–1393. https://doi.org/10.1109/CVPRW.2018.00187

Li, W., Fu, H., Yu, L., Cracknell, A., 2016. Deep Learning Based Oil Palm Tree Detection and Counting for High-Resolution Remote Sensing Images. Remote Sens. 9, 22. https://doi.org/10.3390/rs9010022

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P., 2017. Focal Loss for Dense Object Detection. Proc. IEEE Int. Conf. Comput. Vis. 2017–Octob, 2999–3007. https://doi.org/10.1109/ICCV.2017.324

Liu, T., Im, J., Quackenbush, L.J., 2015. A novel transferable individual tree crown delineation model based on Fishing Net Dragging and boundary classification. ISPRS J. Photogramm. Remote Sens. 110, 34–47. https://doi.org/10.1016/j.isprsjprs.2015.10.002

Mizoguchi, T., Ishii, A., Nakamura, H., Inoue, T., Takamatsu, H., 2017. Lidar-based individual tree species classification using convolutional neural network 103320O. https://doi.org/10.1117/12.2270123

Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017. PointNet: Deep learning on point sets for 3D classification and segmentation. Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017 2017–January, 77–85. https://doi.org/10.1109/CVPR.2017.16

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Nips 91–99. https://doi.org/10.1109/TPAMI.2016.2577031

Uijlings, J.R.R., Van De Sande, K.E.A., Gevers, T., Smeulders, A.W.M., 2013. Selective search for object recognition. Int. J. Comput. Vis. 104, 154–171. https://doi.org/10.1007/s11263-013-0620-5

Vogeler, J.C., Yang, Z., Cohen, W.B., 2016. Mapping post-fire habitat characteristics through the fusion of remote sensing tools. Remote Sens. Environ. 173, 294–303. https://doi.org/10.1016/j.rse.2015.08.011

Weinmann, M., Weinmann, M., Mallet, C., Brédif, M., 2017. A classification-segmentation framework for the detection of individual trees in dense MMS point cloud data acquired in

urban areas. Remote Sens. 9. https://doi.org/10.3390/rs903277

Weinstein, B.G., 2018. A computer vision for animal ecology. J. Anim. Ecol. 87, 533–545. https://doi.org/10.1111/1365-2656.12780

Weinstein, E. White. 2019. weecology/DeepForest: Submission (Version 0.1). Zenodo. http://doi.org/10.5281/zenodo.2538144

Wu, B., Yu, B., Wu, Q., Huang, Y., Chen, Z., Wu, J., 2016. Individual tree crown delineation using localized contour tree method and airborne LiDAR data in coniferous forests. Int. J. Appl. Earth Obs. Geoinf. 52, 82–94. https://doi.org/10.1016/j.jag.2016.06.003

Wu, H., Prasad, S., 2018. Semi-Supervised Deep Learning Using Pseudo Labels for Hyperspectral Image Classification. IEEE Trans. Image Process. 27, 1259–1270. https://doi.org/10.1109/TIP.2017.2772836

Wulder, M.A., Dymond, C.C., White, J.C., Leckie, D.G., Carroll, A.L., 2006. Surveying mountain pine beetle damage of forests: A review of remote sensing opportunities. For. Ecol. Manage. 221, 27–41. https://doi.org/10.1088/0953-4075/23/20/022

Yin, D., Wang, L., 2016. How to assess the accuracy of the individual tree-based forest inventory derived from remotely sensed data: a review. Int. J. Remote Sens. 37, 4521–4553. https://doi.org/10.1080/01431161.2016.1214302

Zhou, Y., Tuzel, O., 2017. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. https://doi.org/1711.06396v1

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geosci. Remote Sens. Mag. 5, 8–36. https://doi.org/10.1109/MGRS.2017.2762307

394

395

9. Supplementary Information

Due to the relatively few number of hand-annotated trees used to create the full model, it is important to have a sense for the sensitivity of the data to the training/validation split. For the hand-annotated trees, 95% of the data was used to train the model, and 5% was used for validation. We performed 10-fold cross validation to determine whether the results were robust to sampling artifacts. We found that metrics varied very little (Table S1). We conclude that sampling have minimal effect on our reported results (Table 1).

Table S1. Evaluation statistics from 10-fold cross validation experiment. The 95% training – 5% validation split was randomized before each run.

Metric	Mean	Min	Max
Recall	0.836	0.805	0.867
mAP@50	0.490	0.445	0.511